Biological Chemistry ‘Just Accepted’ Papers

Biological Chemistry ‘Just Accepted’ Papers are papers published online, in advance of appearing in the print journal. They have been peer-reviewed, accepted and are online published in manuscript form, but have not been copy edited, typeset, or proofread. Copy editing may lead to small differences between the Just Accepted version and the final version. There may also be differences in the quality of the graphics. When papers do appear in print, they will be removed from this feature and grouped with other papers in an issue.

Biol Chem ‘Just Accepted’ Papers are citable; the online publication date is indicated on the Table of Contents page, and the article’s Digital Object Identifier (DOI), a unique identifier for intellectual property in the digital environment (e.g., 10.1515/hsz-2011-xxxx), is shown at the top margin of the title page. Once an article is published as **Biol Chem ‘Just Accepted’ Paper** (and before it is published in its final form), it should be cited in other articles by indicating author list, title and DOI.

After a paper is published in **Biol Chem ‘Just Accepted’ Paper** form, it proceeds through the normal production process, which includes copy editing, typesetting and proofreading. The edited paper is then published in its final form in a regular print and online issue of **Biol Chem**. At this time, the **Biol Chem ‘Just Accepted’ Paper** version is replaced on the journal Web site by the final version of the paper with the same DOI as the **Biol Chem ‘Just Accepted’ Paper version**.

Disclaimer

Biol Chem ‘Just Accepted’ Papers have undergone the complete peer-review process. However, none of the additional editorial preparation, which includes copy editing, typesetting and proofreading, has been performed. Therefore, there may be errors in articles published as **Biol Chem ‘Just Accepted’ Papers** that will be corrected in the final print and online version of the Journal. Any use of these articles is subject to the explicit understanding that the papers have not yet gone through the full quality control process prior to advanced publication.

Research Article

An internally quenched peptide as a new model substrate for rhomboid intramembrane proteases

Elena Arutyunova^{1,a}, Zhenze Jiang^{2,a}, Jian Yang³, Ayodeji N. Kulepa¹, Howard S. Young¹, Steven Verhelst^{3,4}, Anthony J. O'Donoghue^{2#}, M. Joanne Lemieux^{1#}

¹Department of Biochemistry, Faculty of Medicine & Dentistry, Membrane Protein Disease Research Group, Department of Biochemistry, University of Alberta, Edmonton T6G 2R3, Alberta, Canada

²Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

³KU Leuven – University of Leuven, Laboratory of Chemical Biology, Department of Cellular & Molecular Medicine, Herestraat 49 Box 802, B-3000 Leuven, Belgium

⁴Leibniz Institute for Analytical Sciences ISAS, Otto-Hahn-Str. 6b, D-44227 Dortmund, Germany

*Corresponding author

e-mail: mlemieux@ualberta.ca; ajodonoghue@ucsd.edu

^aThese authors contributed equally to this work.

Abstract

Rhomboids are ubiquitous intramembrane serine proteases that cleave transmembrane substrates. Their functions include growth factor signalling, mitochondrial homeostasis, and parasite invasion. A recent study revealed that the *E. coli* rhomboid protease EcGlpG is essential for its extraintestinal pathogenic colonization within the gut. Crystal structures of EcGlpG and the *H. influenzae* rhomboid protease HiGlpG have deciphered an active site that is buried within the lipid bilayer but exposed to the aqueous environment via a cavity at the periplasmic face. A lack of physiological transmembrane substrates has hampered progression for understanding their catalytic mechanism and screening inhibitor libraries. To identify a soluble substrate for use in the study of rhomboid proteases, an array of internally quenched peptides were assayed with HiGlpG, EcGlpG and PsAarA from *P. stuartii*. One substrate was identified that was cleaved by all three rhomboid protease, with HiGlpG having the highest cleavage efficiency. Mass spectrometry analysis determined that all enzymes hydrolyze this substrate between norvaline and tryptophan. Kinetic analysis in both detergent and bicellular systems demonstrated this substrate can be cleaved in solution and in the lipid environment. The substrate was subsequently used to screen a panel of benzoxazin-4-one inhibitors to validate its use in inhibitor discovery.

Keywords: serine protease, intramembrane protease, rhomboid protease, GlpG, kinetics, membrane protein, peptide, bicelles, FRET.

Introduction

Proteases play essential roles in cellular processes and disease (Lopez-Otin and Bond, 2008). These enzymes catalyze the irreversible hydrolysis of peptide bonds and therefore their activity needs to be tightly controlled. Some proteases are promiscuous and cleave many substrates whereas others show specificity for particular substrate sequences. The specificity is mainly determined by amino acid side chain interactions of the substrate to the subsite binding pockets of the protease (Hedstrom, 2002; Perona and Craik, 1995). Understanding this substrate specificity becomes more relevant for proteases that are drug targets (Drag and Salvesen, 2010), especially considering the fact that in the current market proteases account for 5-10% of all pharmaceutical targets (Salisbury and Ellman, 2006).

Intramembrane cleaving proteases are membrane embedded enzymes that harbour a catalytic site within a transmembrane protein (Wolfe, 2009). Similar to soluble proteases several classes have been identified based on the architecture of their active site and mechanism of hydrolysis. These include serine, aspartyl, metallo- and glutamyl intramembrane proteases, all of which participate in a variety of cellular processes in health and disease (Dusterhoft *et al.*, 2017a; Golde *et al.*, 2009; Langosch *et al.*, 2015; Manolaridis *et al.*, 2013; Rawson, 2013).

Kinetic analysis of individual proteases *in vitro* is an important tool to provide substantial insight into specificity and structural determinants required for it. However, the inherent hydrophobic nature of intramembrane proteases comes with particular challenges when *in vitro* work is undertaken. While most membrane proteins retain their stability in detergent solutions, there are many examples where they become unstable when removed from the lipid environment by detergent extraction (Gonzalez Flecha, 2017). In addition, some membrane proteins have preferences for special detergents or lipids. This issue is compounded by the use of substrates that are also transmembrane proteins, therefore absence of activity can be due to instability of the enzyme or the substrate.

Rhomboids are a family of serine intramembrane proteases that are found in all kingdoms (Urban and Dickey, 2011). These enzymes play a role in cell development and microbial pathogenesis (Kateete *et al.*, 2012; Lee *et al.*, 2001; Mesak *et al.*, 2004; Rather *et al.*, 1999; Stevenson *et al.*, 2007; Urban *et al.*, 2001). Recently, the *E. coli* rhomboid protease has been shown to be essential for gut colonization (Russell *et al.*, 2017).

Our understanding of bacterial rhomboid function at the molecular level has been aided by structural studies (Lemieux *et al.*, 2007; Strisovsky, 2013; Wang *et al.*, 2006b) and kinetic studies (Arutyunova *et al.*, 2014; Arutyunova *et al.*, 2016; Dickey *et al.*, 2013). While many substrates are known for eukaryotic rhomboids (Dusterhoft *et al.*, 2017b), only the PsTatA protein from *Providencia stuartii* (Stevenson *et al.*, 2007) has been found to be a physiological prokaryotic rhomboid substrate. Bacterial rhomboid proteases can cleave a variety of eukaryotic substrates, such as Spitz, Gurken, and Keren (Urban *et al.*, 2002) and artificial transmembrane substrates (Erez and Bibi, 2009; Maegawa *et al.*, 2005). The ability of rhomboid proteases to also cleave water-soluble substrates as casein is an apparent paradox, however this substrate has been used to detect rhomboid activity (Lazareno-Saez *et al.*, 2013; Wang *et al.*, 2006; Xue and Ha, 2012).

To identify a soluble fluorescent substrate for activity assessment of bacterial rhomboid proteases, we screened seven commercially available internally quenched substrates, (Table 1S), with recombinant rhomboid proteases from *E. coli* (EcGlpG), *Haemophilus influenza* (HiGlpG) and *P. stuartii* (PsAarA). These enzymes are topologically different (Figure 1), yet they all preferentially hydrolyzed a single substrate that was originally designed for cleavage by human stromelysin 1 (Nagase *et al.*, 1994). Using the three rhomboid proteases, we determined the catalytic efficiencies for this peptide in detergent and bicelles environment and screened a selection of benzoxazin-4-one compounds for inhibition of enzyme activity. We show that this soluble reporter substrate represents a valuable tool for rapid and simple biochemical characterization of bacterial rhomboid proteases.

Results

Identification of internally quenched peptide substrate

Three bacterial rhomboid proteases, the *E. coli* GlpG EcGlpG, the *H. influenza* rhomboid protease HiGlpG, and the *P. stuartii* rhomboid protease PsAarA, were expressed in *E. coli* and purified in the detergent n-Dodecyl-beta-Maltoside (DDM) (Arutyunova *et al.*, 2014). All enzymes were used directly after Immobilized Metal Affinity Chromatography (IMAC) to ensure the maximal activity since it is known that additional steps as Size-Exclusion Chromatography (SEC) can remove endogenously bound lipid molecules, which influence membrane protein stability and activity (Lemieux *et al.*, 2002). The enzymes were assayed

with a selection of internally quenched fluorescent substrates that have been previously used by our group to detect and biochemically characterized proteases in the gut of the freshwater planaria, *Schmidtea mediterranea* (Goupil *et al.*, 2016) and in the secretions of a human skin commensal yeast, *Malassezia glabosa* (Li *et al.*, 2018). These substrates consist of 7-mer to 10-mer peptide sequences flanked by 7-methoxycoumarin (mca) on the amino terminus and dinitrophenol (dnp) conjugated to lysine on the carboxyl-terminus. HiGlpG rapidly cleaved mca-RPKPYAvWM-K(dnp), where lowercase v corresponds to norvaline, a non-natural amino acid (Figure 2). In addition, HiGlpG slowly hydrolyzed mca-GKPILFFRL-K(dnp) and mca-RPPGFASF-K(dnp) however, the activity was negligible when compared to mca-RPKPYAvWM-K(dnp). EcGlpG and PsAarA also cleaved the mca-RPKPYAvWM-K(dnp) peptide but the specific activity was 50- and 135-fold lower, respectively, when compared to HiGlpG (Figure 2). In addition, EcGlpG hydrolyzed mca-GKPILFFRL-K(dnp), mca-RPPGFASF-K(dnp) and mca-HQKLVFFA-K(dnp) while PsAarA cleaved all substrates with the exception of mca-EVKMDAEF-K(dnp).

To determine if the proteolytic activity seen with these soluble substrates was from the recombinant rhomboid enzymes or contaminating *E. coli* proteases, we subjected these enzymes to an additional size exclusion purification step. Activity towards the mca-RPKPYAvWM-K(dnp) peptide was retained, however activity towards the other substrates was reduced suggesting minor contamination with *E. coli* proteases in the IMAC purified preparations (Supplementary Figure S1). In addition, we generated a catalytically inactive HiGlpG, with a serine to alanine mutation at position 116. This protein was purified by IMAC under the same conditions as the active HiGlpG but no hydrolysis of mca-RPKPYAvWM-K(dnp) was detected. Taken together, this substrate screen identified a fluorescent reporter substrates that is hydrolyzed by three bacterial rhomboid proteases.

We next used mass spectrometry to determine the cleavage site within the mca-RPKPYAvWM-K(dnp) substrate. We identified a common cleavage product, mca-RPKPYAv, after 4-hour (HiGlpG) and 24-hour (EcGlpG and PsAarA) incubations of rhomboid proteases with the mca-RPKPYAvWM-K(dnp) (Figure 3, Figure S2), indicating the cleavage site is between norvaline and tryptophan. These data show that the three bacterial enzymes have overlapping substrate specificity; however, they differ in the rate at which they hydrolyze the substrate.

Internally quenched (IQ)-peptide cleavage with detergent-solubilized recombinant rhomboid proteases

To compare the rates of cleavage for the substrate, HiGlpG (200nM), EcGlpG and PsAarA (500nM) were assayed with 0.1 to 100 μM of the mca-RPKPYAvWM-K(dnp) substrate. In parallel, assays were performed using the three enzymes reconstituted in 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) bicelles (Figure 4, Table 1). The linear dependence between the concentrations of the peptide used in the assay and the fluorescence emitted was confirmed (Figure S3). For HiGlpG the catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) was 4-fold faster when the enzyme was assayed in DDM compared to bicelles. Under these same conditions, the catalytic efficiencies for EcGlpG and PsAarA was 4-fold and 58-fold slower in DDM compared to bicelles, respectively.

When comparing the activity in bicelles, the catalytic efficiencies for the three enzymes is within 10-fold (2.3×10^{-2} to 2.8×10^{-3} $\mu\text{M min}^{-1}$) however, in DDM the range increases to >1300-fold; from 0.1 $\mu\text{M min}^{-1}$ for HiGlpG to 7.6×10^{-5} $\mu\text{M min}^{-1}$ for PsAarA. These studies clearly show that activity of these enzymes is highly dependent on the lipid environment surrounding the proteases.

Inhibition of rhomboid proteases by benzoxazin-4-ones inhibitors

Proteases are well known to be targets for drug development, however, to screen for inhibitors that are potent and selective for the target of interest, it is essential to develop a robust assay that is compatible with microplate applications. Using the fluorescent mca-RPKPYAvWM-K(dnp) substrate, we screened a selection of heterocyclic benzoxazin-4-ones compounds that have been shown to inhibit human serine proteases such as cathepsin G, chymotrypsin C1, chymase, thrombin, factor X as well as herpes virus proteases (Powers *et al.*, 2002). More recently, these compounds were reported to inhibit rhomboid proteases EcGlpG and PsAarA in a covalent reversible manner (Goel *et al.*, 2018; Yang *et al.*, 2018). Therefore, we pre-incubated the three bacterial rhomboids solubilized in DDM or reconstituted in bicelles with 100 μM of benzoxazin-4-ones compounds (Figure S4) and then activity was assayed with the mca-RPKPYAvWM-K(dnp) substrate. Using 200 nM of HiGlpG, all of these compounds reduced activity by more than 60% and several compounds reduced activity by >95%. In general, the compounds were more potent against this enzyme in DDM compared to bicelles (Figure 5). Using 500 nM EcGlpG, the same compounds reduced activity by 50% or

more with the exception of JY-18 and JY-39 that had no effect on activity when the enzyme was assayed in DDM. JY-20 reduced activity by 90% in DDM and by 99% in bicelles. Finally, 500 nM of AarA was only partially inhibited by these compounds with the notable exception of JY-20 that completely abolished activity in DDM and reduced activity in bicelles by 77%. These studies demonstrated that the mca-RPKPYAvWM-K(dnp) substrate was an ideal substrate for screening inhibitors of rhomboid activity. In addition, we clearly show that the lipid environment plays a role in turnover of a substrate and potency of inhibitors.

Discussion

Fluorogenic peptide substrates have been used extensively to measure the catalytic activity of soluble proteases. These substrates are amenable to a microplate assay format and therefore facilitate the biochemical characterization of the enzyme while also providing a reporter molecule for the discovery of active site inhibitors (Boulware *et al.*, 2010). Biochemical characterization of intramembrane rhomboid proteases has relied on macromolecular protein substrates such as TatA that contain a cleavable transmembrane region. Hydrolysis at this site can be quantified using SDS-PAGE in the presence and absence of inhibitors (Zoll *et al.*, 2014). Our group previously developed a fluorescent protein substrate consisting of CyPet and YPet flanking the TatA transmembrane substrate. This substrate allowed us determine that binding of this substrate occurs with positive cooperativity (Arutyunova *et al.*, 2014).

Several studies have shed light on rhomboid protease preferences for substrate sequence. For PsAarA the most critical subsites were shown to be P4, P1 and P2' (Strisovsky *et al.*, 2009). Position P1 was found to be the most restrictive with the preference for small amino acids such as Ala, Gly, Cys and Ser. When TatA-derived substrates were mutated to more bulky amino acids a drastic decrease in activity both with PsAarA and EcGlpG enzymes was observed (Strisovsky *et al.*, 2009; Zoll *et al.*, 2014). It has been proposed that the residues in P4-P2' sequence recognition motif determine the cleavage rate of rhomboids but not the K_M (Dickey *et al.*, 2013), implying that interaction area of native rhomboid-substrate complex is usually larger than this segment. Indeed we have shown that these three bacterial rhomboid proteases have specificity preferences for TatA including exosite recognition (Arutyunova *et al.*, 2014). Furthermore, when the cleavage motif was introduced to a region outside of the TatA transmembrane region by mutagenesis, cleavage was still observed suggesting that rhomboid proteases may have a variety of soluble protein substrates (Strisovsky *et al.*, 2009).

This knowledge, and a recent publication from Ticha and coworkers demonstrated that fluorogenic transmembrane peptide substrates can be used for rhomboid intramembrane proteases both in detergent micelles and liposomes for high-throughput screening (Ticha *et al.*, 2017), encouraged us to screen soluble fluorescent peptides to discover a reporter substrate that can be readily hydrolyzed by this group of enzymes.

Many of the well characterized soluble serine proteases hydrolyze short tri- and tetra-peptide substrates that have fluorescent reporter molecule in the P1' position (Debela *et al.*, 2006; Harris *et al.*, 2000; Kasperkiewicz *et al.*, 2015). These substrates lack amino acids in the P2' position and therefore are unlikely to be hydrolyzed by the bacterial rhomboids. Aspartic acid protease and metalloproteases generally hydrolyze peptide substrates that have prime and non-prime side amino acids and many internally quenched fluorescent reporter substrates have been developed for these enzymes. These substrates generally consist of a 6- to 10-mer peptide sequence flanked on one terminus by a fluorophore and on the other terminus by a compatible quenching molecule. Hydrolysis of the peptide results in an increase in fluorescent intensity that can be readily monitored in a microplate assay. Using the bacterial rhomboids, we screened a selection of peptide substrates that were originally developed for assaying the metalloproteases ADAM10, MMP3, and endothelin-converting enzyme-1, the aspartic acid proteases, BACE1 and cathepsin D and the cysteine protease, caspase-3. We found that one of these substrates, mca-RPKPYAvWM-K(dnp), was hydrolyzed by all enzymes and cleavage occurred between norvaline in the P1 position and Trp in the P1' position. Interestingly, Strisovsky and colleagues showed that P1 amino acids containing branched aliphatic side chains such as Val, Leu and Ile were not hydrolyzed by PsAarA, while unbranched aliphatic amino acid such as Gly, Ala, and to some degree Met, are preferred (Strisovsky *et al.*, 2009). EcGlpG appears to be even more selective at this position with only P1-Ala and P1-Cys containing substrates being hydrolyzed efficiently. Norvaline is a non-proteinogenic unbranched isomer of valine with a side chain that is intermediate in length to Ala and Met. The preference of rhomboids for this amino acid cannot be evaluated in macromolecular substrates as they rely on ribosomal protein synthesis however, synthetic substrates consisting of non-natural amino acids have been shown to be useful for designing substrates for soluble proteases (Poreba *et al.*, 2017).

Outside of the P1 position, EcGlpG and PsAarA substrate specificity has been thoroughly investigating using the TatA substrate by mutation of the amino acids in the P5 to P2 positions. For these studies cleavage of the mutant TatA proteins was either slower, faster or

unchanged when compared to the wild-type enzyme that contains the sequence TIATA in the P5 to P2 positions. To better understand how the fluorescent substrate is cleaved by the bacterial rhomboids, we compared the P5 to P2 residues (KPYAv) of the fluorescence substrate, to the substrate preference of EcGlpG and PsAarA. When P5-Thr or P3-Ala of TatA was mutated to P5-Lys or P3-Tyr, hydrolysis by EcGlpG was increased (Zoll *et al.*, 2014) while PsAarA activity was unchanged (Strisovsky *et al.*, 2009). In the P4 position, mutation of Ile in the wild-type protein to Pro resulted in increased hydrolysis by EcGlpG but greatly decreased the rate of hydrolysis by PsAarA. In addition, mutation of P2 Thr to Ala did not change the rate of TatA cleavage for either enzyme. Taken together, these data indicate that P5-Lys, P3-Tyr and P2-Thr found in the fluorescent substrate are favorable or neutral amino acids for PsAarA and EcGlpG, however, P4-Pro is much more favorable for EcGlpG.

The prime side substrate specificity of PsAarA has also been investigated using the TatA protein substrates (Strisovsky *et al.*, 2009). TatA contains Ala at P1' and Phe at P2' while the internally quenched substrate consists of Trp at P1' and Met at P2'. When the P1' residue of TatA was mutated to Trp, the rate of hydrolysis by PsAarA was decreased by more than 50% while mutation of Phe to Met in the P2' position had no effect substrate hydrolysis. These data indicate that catalytic efficiency of this enzyme could be improved by designing a fluorescent substrate with Ala in the P1' instead of Trp.

Rhomboids cleave both transmembrane and soluble substrates with loop 5 and helix 5 thought to be responsible for substrate gating, providing access of soluble substrates to the buried active site (Baker *et al.*, 2007; Brooks *et al.*, 2011; Wang and Ha, 2007). Cleavage of transmembrane substrates is a multi-step kinetic scheme where substrate is recognized by the exosite and this initial binding facilitates substrate entry laterally into the active site (Arutyunova *et al.*, 2014; Cho *et al.*, 2016; Dickey *et al.*, 2013; Strisovsky *et al.*, 2009). The presence of an additional binding site (exosite) provides a separate step for discriminating substrates from non-substrates, but at the same time represents a rate-limiting step for catalysis. We anticipate that peptide substrates gain access to the active site from the soluble face of the enzyme via the opening of loop 5, which would be exposed in both detergent and bicelles environments. This would eliminate the initial binding step laterally from the lipid bilayer partition, resulting in a faster cleavage rate.

Detergent micelles remain the most common membrane mimetic for protein extraction, refolding, crystallization and functional studies however it is widely known that secondary

and tertiary structures, stability, folding and aggregation of membrane proteins are all dependant on the membrane environment. Determining the optimal membrane-mimicking system is essential for functional assessment of rhomboid proteases. When detergents such as CHAPS or CHAPSO are mixed with short chain lipids such as DMPC or DHPC in the correct ratio and at the correct temperature, bilayered discoidal structures known as bicelles are formed (Durr *et al.*, 2012). Bicelles are applicable for structural and functional studies of membrane proteins and have been used to assess intramembrane protease activity and structure (Sherratt *et al.*, 2009; Vinothkumar, 2011). Therefore we compared enzyme activity in bicelle-reconstituted rhomboid enzymes to the same enzyme in detergent micelles using the fluorescent mca-RPKPYAvWM-K(dnp) substrate and found the EcGlpG and PsAarA activity, but not HiGlpG, was highly dependent on the lipid environment. Bicelles may provide structural stability for EcGlpG and PsAarA as these enzyme have either a cytoplasmic domain (EcGlpG) or an additional predicted 7th transmembrane segment (PsAarA) that is absent in HiGlpG (Panigrahi *et al.*, 2016).

To evaluate the peptide substrate for use in inhibitor screening, we evaluated a selection of serine protease antagonists that have been shown to inhibit EcGlpG and YqgP rhomboids (Goel *et al.*, 2018; Yang *et al.*, 2018). Assays were performed in detergent micelles and bicelles systems and we show that the lipid environment also affects the potency of these compounds. In general HiGlpG was sensitive to most compounds whereas PsAarA was the most tolerant. A notable exception was compound JY-20 that potently inhibited all three rhomboids.

In summary, we have established the use of an internally quenched fluorescent peptide as a soluble substrate peptide to assess functionality of bacterial rhomboid proteases. We have demonstrated the effect of a bicelle system on rhomboid protease activity and conclude that the increase in activity for bicelle-reconstituted proteases solely reflects the stabilizing changes in the active site of the enzyme and does not depend on accessibility of the substrate. This peptide substrate represents a convenient tool to evaluate different membrane-mimicking systems to identify conditions for optimal cleavage of soluble substrates. In addition, this substrate can be used for high-throughput screening of inhibitors for rhomboid proteases.

Materials and methods

Isolation of recombinant bacterial rhomboid proteases

Purification of EcGlpG, HiGlpG, and PsAarA was conducted as previously described (Arutyunova *et al.*, 2014). In brief, the pBadMycHisA vector bearing rhomboid gene was transformed into TOP10 cells and induced with 0.002% arabinose. Protein expression occurred at 24°C for 8 h in LB media. Cells were lysed in 50 mM Tris-HCl, pH 8, 150 mM NaCl buffer and membranes were isolated by ultracentrifugation. The protein was purified from DDM solubilized membranes using nickel immobilized metal affinity chromatography (IMAC). The protein samples were eluted with imidazole buffer exchanged and concentrated using 30 000 MwCO concentrators (Millipore, USA), flash-frozen and stored at -80°C. Proteins were also purified using IMAC followed by size-exclusion chromatography on a Superdex 200 (16/60) (GE Health, Canada) using 50mM Tris-HCl, pH 8, 150mM NaCl, 5% glycerol, 0.1% DDM. Fractions were concentrated as outlined above.

Bicelle preparation

Bicelles were prepared by mixing DMPC and CHAPSO in the molar ratio of 3 to 1 by performing cycles of brief heating (~40-50°C), cooling on ice, vortexing and freezing. The final bicelle concentration was chosen to be in the range between 25%-35%.

Internally-quenched fluorescent substrate screening.

All substrates we purchased from CPC Scientific, Sunnyvale, CA, USA. For screening 2.7 μ M of EcGlpG, 2.3 μ M of HiGlpG and 9.6 μ M of PsAarA were incubated with 5 μ M of internally quenched substrates in the buffer containing: 20 mM Citrate-Phosphate, pH 6.0, 150 mM NaCl, 2 mM DTT, 20% Glycerol and 0.1% DDM. Fluorescent intensity was measured at Ex/Em: 320nm/400nm using a Synergy HTX microplate reader (BioTek) at room temperature. The specific activity was calculated by using initial velocity. For determining cleavage site, proteases were incubated with 5 μ M of mca-RPKPYAvWM-K(dnp) for 6 hours (HiGlpG) or 24 hours (EcGlpG and AarA) at room temperature and the reaction was quenched by the addition of 8 M GuHCl. The samples were desalted using a C18 tips (Ramin), eluted in 80 % acetonitrile and dried down in a vacuum centrifuge prior to mass spectrometry analysis.

Liquid chromatography and mass spectrometry

Q Exactive Mass Spectrometer (Thermo) was equipped with an Ultimate 3000 HPLC. Peptides were separated by reverse phase chromatography on a C18 column (1.7 μm bead size, 75 μm x 25 cm, heated to 65°C) at a flow rate of 300 nl min^{-1} using a 20 min linear gradient from 5% B to 50% B, with solvent A: 0.1% formic acid in water and solvent B: 0.1% formic acid in acetonitrile. Survey scans were recorded over a 150–2000 m/z range at 70 000 resolution at 200 m/z (AGC target 1×10^6 , 75 ms maximum injection time). MS/MS was performed in data-dependent acquisition mode targeting all possible cleavages product of mca-RPKPYAvWM-K(dnp) with HCD fragmentation (30 normalized collision energy) on the 5 most intense precursor ions at 17500 resolution at 200 m/z (AGC target 5×10^4 , 120 ms maximum injection time, dynamic exclusion 15s).

IQ-peptide-based protease kinetic assay

The enzymatic assay was performed as previously described (Arutyunova *et al.*, 2014). Briefly, EcGlpG (500 nM), HiGlpG (200 nM) and PsAarA (500 nM) were incubated with 0.1–100 μM of mca-RPKPYAvWM-K(dnp) peptide at 37°C for 2 hours in activity buffer containing: 50mM MES pH 6.0, 150 mM NaCl, 20% glycerol and 0.1% DDM for detergent-solubilized enzymes and in the same buffer but without the detergent for bicelle-reconstituted enzymes. It was verified that the concentration of proteases used in the assay gave a linear dependence of amount of generated product with time that was chosen. The FRET-based cleavage assay was conducted in a fluorescence multiwell plate reader (SynergyMx, BioTek). Upon excitation at 320 nm, the emission intensity at 400 nm was recorded at 3 min intervals for 2 hours and plotted with GraphPad PRISM software. The initial velocities were defined as the slopes of a graph of generated product as a function of time. The plots of the generated product against the corresponding substrate concentration were fitted with Michaelis–Menten equation. Catalytic parameters were extrapolated with GraphPad PRISM. Additionally, it was verified that the proportionality between the fluorescence emitted and the amount of the substrate used in the assay was linear (Figure S4). The RFU to μM conversion factor was obtained from the plot represented on Figure S4.

Inhibition assay

Inhibitors were dissolved in DMSO at the stock concentration of 20 mM. For each enzyme, 500 nM, in either DDM or bicelles, was incubated with 100 μM of inhibitor for 40 min in

activity buffer. DMSO (2% v/v) was used as negative control. The reaction was started with the substrate and fluorescence was monitored in a fluorescence multiwell plate reader (SynergyMx, BioTek) for 2 h at 37°C.

Acknowledgements

Funding: M.J.L. gratefully acknowledges supported by grant MOP-93557 from the Canadian Institutes of Health Research, Neuroscience and Mental Health Institute (Brad Mates Foundation), Parkinson Alberta, and the Department of Biochemistry, University of Alberta. M.J.L. was supported in part by funding from Alberta Innovates Health Solutions and the Parkinson's Society of Canada. J.Y. acknowledges funding by a CSC scholarship. S.V. was supported in part by the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen, the Senatsverwaltung für Wirtschaft, Technologie und Forschung des Landes Berlin and the Bundesministerium für Bildung und Forschung. A.J.O. gratefully acknowledges financial support from UC San Diego Skaggs School of Pharmacy and Pharmaceutical Science.

Conflict of interest statement

The authors declare that they have no conflicts of interest concerning the contents of this article.

References

- Arutyunova, E., Panwar, P., Skiba, P.M., Gale, N., Mak, M.W., and Lemieux, M.J. (2014). Allosteric regulation of rhomboid intramembrane proteolysis. *EMBO J* 33, 1869-1881.
- Arutyunova, E., Smithers, C.C., Corradi, V., Espiritu, A.C., Young, H.S., Tieleman, D.P., and Lemieux, M.J. (2016). Probing catalytic rate enhancement during intramembrane proteolysis. *Biol Chem* 397, 907-919.
- Baker, R.P., Young, K., Feng, L., Shi, Y., and Urban, S. (2007). Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc Natl Acad Sci USA* 104, 8257-8262.
- Boulware, K.T., Jabaiah, A., and Daugherty, P.S. (2010). Evolutionary optimization of peptide substrates for proteases that exhibit rapid hydrolysis kinetics. *Biotechnol Bioeng* 106, 339-346.
- Brooks, C.L., Lazareno-Saez, C., Lamoureux, J.S., Mak, M.W., and Lemieux, M.J. (2011). Insights into substrate gating in *H. influenzae* rhomboid. *J Mol Biol* 407, 687-697.
- Cho, S., Dickey, S.W., and Urban, S. (2016). Crystal structures and inhibition kinetics reveal a two-stage catalytic mechanism with drug design implications for rhomboid proteolysis. *Mol Cell* 61, 329-340.
- Debela, M., Magdolen, V., Schechter, N., Valachova, M., Lottspeich, F., Craik, C.S., Choe, Y., Bode, W., and Goettig, P. (2006). Specificity profiling of seven human tissue kallikreins reveals individual subsite preferences. *J Biol Chem* 281, 25678-25688.
- Dickey, S.W., Baker, R.P., Cho, S., and Urban, S. (2013). Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* 155, 1270-1281.
- Drag, M., and Salvesen, G.S. (2010). Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov* 9, 690-701.
- Durr, U.H., Gildenberg, M., and Ramamoorthy, A. (2012). The magic of bicelles lights up membrane protein structure. *Chem Rev* 112, 6054-6074.
- Dusterhoft, S., Kunzel, U., and Freeman, M. (2017a). Rhomboid proteases in human disease: Mechanisms and future prospects. *Biochim Biophys Acta*
- Dusterhoft, S., Kunzel, U., and Freeman, M. (2017b). Rhomboid proteases in human disease: Mechanisms and future prospects. *Biochim Biophys Acta* 1864, 2200-2209.
- Erez, E., and Bibi, E. (2009). Cleavage of a multispinning membrane protein by an intramembrane serine protease. *Biochemistry* 48, 12314-12322.
- Goel, P., Jumpertz, T., Ticha, A., Ogorek, I., Mikles, D.C., Hubalek, M., Pietrzik, C.U., Strisovsky, K., Schmidt, B., and Weggen, S. (2018). Discovery and validation of 2-styryl substituted benzoxazin-4-ones as a novel scaffold for rhomboid protease inhibitors. *Bioorg Med Chem Lett* 28, 1417-1422.
- Golde, T.E., Wolfe, M.S., and Greenbaum, D.C. (2009). Signal peptide peptidases: a family of intramembrane-cleaving proteases that cleave type 2 transmembrane proteins. *Semin Cell Dev Biol* 20, 225-230.
- Gonzalez Flecha, F.L. (2017). Kinetic stability of membrane proteins. *Biophys Rev* 9, 563-572.

- Goupil, L.S., Ivry, S.L., Hsieh, I., Suzuki, B.M., Craik, C.S., O'Donoghue, A.J., and McKerrow, J.H. (2016). Cysteine and aspartyl proteases contribute to protein digestion in the gut of freshwater planaria. *PLoS Negl Trop Dis* 10, e0004893.
- Harris, J.L., Backes, B.J., Leonetti, F., Mahrus, S., Ellman, J.A., and Craik, C.S. (2000). Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc Natl Acad Sci USA* 97, 7754-7759.
- Hedstrom, L. (2002). An overview of serine proteases. *Curr Protoc Protein Sci* Chapter 21, Unit 21 10.
- Kasperkiewicz, P., Poreba, M., Snipas, S.J., Lin, S.J., Kirchhofer, D., Salvesen, G.S., and Drag, M. (2015). Design of a selective substrate and activity based probe for human neutrophil serine protease 4. *PLoS One* 10, e0132818.
- Kateete, D.P., Katabazi, F.A., Okeng, A., Okee, M., Musinguzi, C., Asiimwe, B.B., Kyobe, S., Asiimwe, J., Boom, W.H., and Joloba, M.L. (2012). Rhomboids of Mycobacteria: characterization using an *aarA* mutant of *Providencia stuartii* and gene deletion in *Mycobacterium smegmatis*. *PLoS One* 7, e45741.
- Langosch, D., Scharnagl, C., Steiner, H., and Lemberg, M.K. (2015). Understanding intramembrane proteolysis: from protein dynamics to reaction kinetics. *Trends Biochem Sci* 40, 318-327.
- Lazareno-Saez, C., Arutyunova, E., Coquelle, N., and Lemieux, M.J. (2013). Domain swapping in the cytoplasmic domain of the *Escherichia coli* rhomboid protease. *J Mol Biol* 425, 1127-1142.
- Lee, J.R., Urban, S., Garvey, C.F., and Freeman, M. (2001). Regulated intracellular ligand transport and proteolysis control EGF signal activation in *Drosophila*. *Cell* 107, 161-171.
- Lemieux, M.J., Fischer, S.J., Cherney, M.M., Bateman, K.S., and James, M.N. (2007). The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc Natl Acad Sci USA* 104, 750-754.
- Lemieux, M.J., Reithmeier, R.A., and Wang, D.N. (2002). Importance of detergent and phospholipid in the crystallization of the human erythrocyte anion-exchanger membrane domain. *J Struct Biol* 137, 322-332.
- Li, H., Goh, B.N., Teh, W.K., Jiang, Z., Goh, J.P.Z., Goh, A., Wu, G., Hoon, S.S., Raida, M., Camattari, A., *et al.* (2018). Skin commensal *Malassezia globosa* secreted protease attenuates *Staphylococcus aureus* biofilm formation. *J Invest Dermatol* 138, 1137-1145.
- Lopez-Otin, C., and Bond, J.S. (2008). Proteases: multifunctional enzymes in life and disease. *J Biol Chem* 283, 30433-30437.
- Maegawa, S., Ito, K., and Akiyama, Y. (2005). Proteolytic action of GlpG, a rhomboid protease in the *Escherichia coli* cytoplasmic membrane. *Biochemistry* 44, 13543-13552.
- Manolaridis, I., Kulkarni, K., Dodd, R.B., Ogasawara, S., Zhang, Z., Bineva, G., O'Reilly, N., Hanrahan, S.J., Thompson, A.J., Cronin, N., *et al.* (2013). Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. *Nature* 504, 301-305.
- Mesak, L.R., Mesak, F.M., and Dahl, M.K. (2004). Expression of a novel gene, *gluP*, is essential for normal *Bacillus subtilis* cell division and contributes to glucose export. *BMC Microbiol* 4, 13.

- Nagase, H., Fields, C.G., and Fields, G.B. (1994). Design and characterization of a fluorogenic substrate selectively hydrolyzed by stromelysin 1 (matrix metalloproteinase-3). *J Biol Chem* 269, 20952-20957.
- Panigrahi, R., Arutyunova, E., Panwar, P., Gimpl, K., Keller, S., and Lemieux, M.J. (2016). Reversible Unfolding of rhomboid intramembrane proteases. *Biophys J* 110, 1379-1390.
- Perona, J.J., and Craik, C.S. (1995). Structural basis of substrate specificity in the serine proteases. *Protein Sci* 4, 337-360.
- Poreba, M., Salvesen, G.S., and Drag, M. (2017). Synthesis of a HyCoSuL peptide substrate library to dissect protease substrate specificity. *Nat Protoc* 12, 2189-2214.
- Powers, J.C., Asgian, J.L., Ekici, O.D., and James, K.E. (2002). Irreversible inhibitors of serine, cysteine, and threonine proteases. *Chem Rev* 102, 4639-4750.
- Rather, P.N., Ding, X., Baca-DeLancey, R.R., and Siddiqui, S. (1999). *Providencia stuartii* genes activated by cell-to-cell signaling and identification of a gene required for production or activity of an extracellular factor. *J Bacteriol* 181, 7185-7191.
- Rawson, R.B. (2013). The site-2 protease. *Biochim Biophys Acta* 1828, 2801-2807.
- Russell, C.W., Richards, A.C., Chang, A.S., and Mulvey, M.A. (2017). The rhomboid protease GlpG promotes the persistence of extraintestinal pathogenic *Escherichia coli* within the gut. *Infection and immunity* 85.
- Salisbury, C.M., and Ellman, J.A. (2006). Rapid identification of potent nonpeptidic serine protease inhibitors. *Chembiochem* 7, 1034-1037.
- Sherratt, A.R., Braganza, M.V., Nguyen, E., Ducat, T., and Goto, N.K. (2009). Insights into the effect of detergents on the full-length rhomboid protease from *Pseudomonas aeruginosa* and its cytosolic domain. *Biochim Biophys Acta* 1788, 2444-2453.
- Stevenson, L.G., Strisovsky, K., Clemmer, K.M., Bhatt, S., Freeman, M., and Rather, P.N. (2007). Rhomboid protease AarA mediates quorum-sensing in *Providencia stuartii* by activating TatA of the twin-arginine translocase. *Proc Natl Acad Sci USA* 104, 1003-1008.
- Strisovsky, K. (2013). Structural and mechanistic principles of intramembrane proteolysis--lessons from rhomboids. *FEBS J* 280, 1579-1603.
- Strisovsky, K., Sharpe, H.J., and Freeman, M. (2009). Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* 36, 1048-1059.
- Ticha, A., Stanchev, S., Skerle, J., Began, J., Ingr, M., Svehlova, K., Polovinkin, L., Ruzicka, M., Bednarova, L., Hadravova, R., *et al.* (2017). Sensitive versatile fluorogenic transmembrane peptide substrates for rhomboid intramembrane proteases. *J Biol Chem* 292, 2703-2713.
- Urban, S., and Dickey, S.W. (2011). The rhomboid protease family: a decade of progress on function and mechanism. *Genome Biol* 12, 231.
- Urban, S., Lee, J.R., and Freeman, M. (2001). *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* 107, 173-182.

- Urban, S., Schlieper, D., and Freeman, M. (2002). Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Curr Biol* *12*, 1507-1512.
- Vinothkumar, K.R. (2011). [Structure of rhomboid protease in a lipid environment. J Mol Biol](#) *407*, 232-247.
- Wang, Y., and Ha, Y. (2007). Open-cap conformation of intramembrane protease GlpG. *Proc Natl Acad Sci USA* *104*, 2098-2102.
- Wang, Y., Zhang, Y., and Ha, Y. (2006). Crystal structure of a rhomboid family intramembrane protease. *Nature* *444*, 1-5.
- [Wolfe, M.S. \(2009\). Intramembrane-cleaving proteases. J Biol Chem](#) *284*, 13969-13973.
- Xue, Y., and Ha, Y. (2012). Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J Biol Chem* *287*, 3099-3107.
- Yang, J., Barniol-Xicota, M., Nguyen, M.T.N., Ticha, A., Strisovsky, K., and Verhelst, S.H.L. (2018). Benzoxazin-4-ones as novel, easily accessible inhibitors for rhomboid proteases. *Bioorg Med Chem Lett* *28*, 1423-1427.
- Zoll, S., Stanchev, S., Began, J., Skerle, J., Lepsik, M., Peclinovska, L., Majer, P., and Strisovsky, K. (2014). Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J* *33*, 2408-2421.

Tables and figures

Table 1 Catalytic parameters of EcGlpG, HiGlpG and PsAarA with the internally quenched mca-RPKPYAvWM-K(dnp) substrate.

	HiGlpG		EcGlpG		PsAarA	
	DDM	Bicelles	DDM	Bicelles	DDM	Bicelles
K_M (μM)	20 $\pm$ 2	58 $\pm$ 9	22 $\pm$ 3	17 $\pm$ 4	71 $\pm$ 17	25 $\pm$ 5
k_{cat} (min^{-1})	2.1 $\pm$ 0.2	1.35 $\pm$ 0.1	0.014 $\pm$ 6 $\times 10^{-4}$	0.048 $\pm$ 0.004	0.0054 $\pm$ 0.001	0.112 $\pm$ 0.008
k_{cat}/K_M ($\mu\text{M min}^{-1}$)	0.1 $\pm$ 0.02	0.023 $\pm$ 0.005	6.4 $\times 10^{-4}$ $\pm 1.1 \times 10^{-4}$	2.8 $\times 10^{-3}$ $\pm 0.84 \times 10^{-3}$	7.6 $\times 10^{-5}$ $\pm 3 \times 10^{-5}$	4.4 $\times 10^{-3}$ $\pm 0.9 \times 10^{-3}$

Hydrolysis rates for the mca-RPKPYAvWM-K(dnp) substrate were measured with the three rhomboid proteases, in DDM or in a bicelles environment. The data is represented as mean ($n=3$) $\pm$ SD.

Figure 1 Schematic representation of topologies of three prokaryotic rhomboid proteases, HiGlpG, EcGlpG and PsAarA.

Rhomboid protease quality control

Figure 2 Fluorescent substrate screening of rhomboid proteases.

All substrates are flanked with mca fluorophore on N-terminus and Lys(dnp) quencher on the C-terminus.

Figure 3 Mass spectrometry determined the cleavage site of mca-RPKPYAvWM-K(dnp) substrate is between norvaline and tryptophan.

Peak 1: mca-RPKPYAvWM-K(dnp) K(dnp)-NH₂; Peak 2: mca-RPKPYAv.

Figure 4 Saturation curves of three bacterial rhomboids in DDM and bicelles with mca-RPKPYAvWM-K(dnp) peptide as substrate.

(A) Saturation curves of HiGlpG in DDM and bicelles with the peptide. (B) Saturation curves of EcGlpG in DDM and bicelles with the peptide. (C) Saturation curves of PsAarA in DDM and bicelles with the peptide.

A **B** **C**

Figure 5 The residual activities of HiGlpG, EcGlpG and PsAarA in DDM and bicelles treated with benzoxazin-4-ones inhibitors.

Data is represented as mean ($n=3$) $\pm$ SE.

Supplementary material

Figure S1 A. The S116A inactive mutant of HiGlpG is unable to cleave the internally quenched peptide, RPKPYAvWM. B,C,D. Assessment of the internally quenched peptide substrate cleavage reveals that compared to post immobilized metal affinity chromatography

(IMAC) there is a loss of hydrolysis after size exclusion chromatography (SEC). RFU: relative fluorescence units. The data presented as mean ($n=3$) $\pm$ SE.

Figure S2 Mass spectrometry was used to identify the cleavage site within the RPKPYAvWM substrate. After 4 hours incubation of HiGlpG with the substrate, we discovered that the full length Mca-RPKPYAvWMK(DNP) substrate that eluted between minute 20 and 21 was no longer detectable and a single cleavage product corresponding to RPKPYAv was found to elute at minute 14.

Figure S3 The proportionality between the concentration of substrate used in the kinetic assay and the fluorescence emitted.

Figure S4 Chemical structures of benoxazin-4-ones inhibitors used in this study.